Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7749324, 11 pages
https://doi.org/10.1155/2022/7749324

Research Article

Movement Evaluation Algorithm-Based Form Tracking Technology and Optimal Control of Limbs for Dancers

Jia Chen [1] **and Luxi Chen** [2]

[1]*College of Music and Dance, Yulin Normal University, Yulin, Guangxi 537000, China*
[2]*College of History Culture and Tourism, Yulin Normal University, Yulin, Guangxi 537000, China*

Correspondence should be addressed to Luxi Chen; chenluxi12345678@ylu.edu.cn

Received 2 August 2022; Revised 6 September 2022; Accepted 17 September 2022; Published 11 October 2022

Academic Editor: Gengxin Sun

This study designs an optimal control model of dance personnel from tracking technology and limb control based on an action evaluation algorithm by constructing a human action evaluation algorithm model and conducting an in-depth study of dance personnel from tracking technology and limb control. This study proposes an OpenPose method based on pose flow optimization to address the false detection of vital human points and misconnection between critical issues in traditional OpenPose-based human pose estimation. The human pose estimation results of OpenPose are optimized by using the human pose flow information in the image sequence. This makes up for the shortcomings of traditional OpenPose that ignores the interframe image information. In this study, we analyze the experimental data of action evaluation, define a set of formulas to evaluate the action after summarizing the distribution pattern of DTW difference sample points from 8 angles, and design an action evaluation system to demonstrate the rationality of this action evaluation method. Since the bases and factors in the evaluation formula are constantly recalculated as the action changes, which increases the complexity of the evaluation method, the following work is to improve the parameters of the activity evaluation formula, so that the evaluation method has better efficiency and adaptability. To enhance the effect of action recognition, this study uses the Kinect sensor to obtain the 3D coordinates of 20 skeletal joints of the human body. It uses the relative distance and angle sequence of the joints as the feature parameters according to the characteristics of human posture. In static pose recognition, the feature vector's sample set is trained, and the KNN algorithm is used as a classifier to recognize the pose.

1. Introduction

Human movement is the behavior produced by the whole body or part of the human limbs to convey information, which expresses feelings and meets various needs of people in life; evaluation is the conclusion after judging and analyzing a thing or a person [1]. People often evaluate various human body movements in daily life, such as coaches teaching fitness movements, judges in diving, competitive gymnastics, and other competitions when scoring and doctors rehabilitating patients [2]. At present, the evaluation of human body movements is mainly done by human beings, and there are two problems when human beings evaluate human body movements; one is that they cannot be assessed objectively; for example, when judges score in competitions, they are easily influenced by subjective factors and cannot be judged impartially; second, the evaluation results that only rely on the human eyes are not accurate enough and often overlook some details. Therefore, there is a need for technology to realize the evaluation from human-to-human action to the machine. In recent years, with the advancement of computer hardware and software technology, the research on human movement has become increasingly in-depth through various sensor technologies and artificial intelligence algorithms, and human movement analysis has now become a research hotspot [3]. As new research in human movement analysis, human movement evaluation is also gaining more and more attention, and this technology has a

wide range of application prospects. It can be applied to medical rehabilitation, intelligent fitness, physical games, and other fields in the future.

Current approaches to human motion modeling fall into two main following categories: marker point-based motion capture and vision-based motion capture. Marker point-based motion capture requires different types of sensors installed on the human body [4]. Although these methods can model human motion in real-time accurately and robustly, they need a specific indoor environment and clothing; so, they are more costly. Vision-based methods use single or multiple cameras to capture and model human motion through computer vision and graphics [5]. These techniques overcome the limitations of wearing specific clothing and specific acquisition environments to model more flexible, diverse, and natural human motion outdoors. Stable capture of highly accurate human activity is the goal of motion modeling, which can be achieved by tracking human feature points over time, i.e., attain human motion tracking. Human morphology estimation is to locate each key point of the human body in the 2D image to determine the human morphology, and the position of each key point is based on the coordinates of the 2D image [6]. The motion modeling approach proposed in subsequent sections of this study revolves around 2D human morphology estimation. The rapid development of artificial intelligence and computer vision technology has been widely used in security, criminal investigation, sports, entertainment, rehabilitation therapy, and other fields. Human action recognition and analysis, an important research direction of computer vision, has received increasing attention [7]. At this stage, numerous studies are on human form estimation and analysis; however, most of them aim to perform human action recognition. However, for sports with strict specifications and routines, such as Taijiquan, gymnastics, and dance, it is not enough to have the techniques for recognition. It is also necessary to evaluate their consistency with standard movements for guidance.

On the other hand, most of the current methods for human movement analysis are based only on unimodal information to characterize human morphology, such as color images, skeletal information, and depth images. The information source of action characterization is single, which is challenging to characterize the human activity process effectively and accurately [8] and has considerable limitations. In contrast, multimodal action information can reflect the complementarity of different modal information and improve the accuracy of human action characterization. In addition, the characteristics of the human visual attention mechanism and the contextual relevance of human action sequences are often ignored in human action recognition based on deep learning, which makes it challenging to analyze human actions intelligently and effectively [9]. Human action analysis based on multimodal information is a hot research topic in computer vision. The human action process is very complex, and how to effectively characterize human action sequences based on image information and skeletal information is an important research content. At the same time, how to build an intelligent and efficient human action analysis model for multimodal human action features is a pressing challenge. This study analyzes and evaluates multimodal human actions. Based on the image information, we propose the OpenPose method for morphological flow optimization to obtain human morphological information [10]. Based on the human morphology information, the deep learning method is selected to offer the morphology tracking technology and optimal limb control of dancers based on an action evaluation algorithm. The two models are analyzed and identified for the spatiotemporal characteristics of human movements using feature fusion and decision fusion strategies, respectively. Based on this, this study develops an intelligent interactive action evaluation system for dancers' morphology, which can evaluate human movements accurately in real-time.

2. Related Works

In computer vision, target tracking is a critical technology that monitors a target object and reasonably predicts its activity trajectory through a detailed analysis of pixel points in video sequence frames to achieve real-time tracking of the target object [11]. There are various approaches to human tracking, such as model-based tracking methods that treat the target as a whole and monitor the position of the human body in each scene while also analyzing the direction, route, and rate of human movement [12]. By taking the line drawing method or the contour method, the target is regarded as a specific geometric shape; for example, in surveillance video or camera photos, oval or square contours are often found in the target location. While tracking human targets, we also encounter interference factors of complex backgrounds such as shadows, occlusions, and textures, which can significantly increase the difficulty of target tracking [13]. In the current research on target tracking, only the target object is moving unilaterally, which is unsuitable for some special situations, such as on the race course or in the military. Contemporary examples of the simultaneous movement of the target object and the camera are also less seen, limiting the application of human target tracking in a broader range of fields.

Unlike research on human action recognition, action detection, and action segmentation, vision-based human action evaluation techniques evaluate the quality of human actions in a video [14]. Some initial progress has been made in the field of human action evaluation. First, RGB images or depth images of the human body are acquired. Then, human skeletal critical point data are obtained from these image data, followed by extracting human action features from the essential skeletal data of the point and then comparing the similarity of the features afterward to obtain the evaluation results [15]. After acquiring the human action image data, it is difficult to evaluate the human action directly through these data because the positioning of the human body in the image space is a complex process. There is a lot of background information in these images that are unrelated to the human action, which needs to be eliminated, and only the human skeleton key point data is retained [16]. OpenPose is a popular open-source library for acquiring 2D skeletal

joints of the human body for single or multiperson motion capture and is highly resistant to interference.

In motion reorientation, early researchers used traditional numerical methods. Qi X first proposed this problem as a spatial-temporal optimization problem for the entire motion, centered on solving it using hand-designed constraints and inverse kinematics [17]. Wu and Huo proposed a hierarchical curve-fitting technique combined with inverse kinematics to solve this problem [18]. Wu et al. used a relay skeleton to migrate the motion between two characters with different levels or geometries. The introduction of learning-based methods effectively solves the above challenges [19]. A two-layer recurrent neural network structure was proposed by Yan et al. Due to the lack of good data pairs for motion redirection, an unsupervised approach is proposed to train the model [20]. The process is to attach a layer of forwarding kinematic layer after the recurrent neural network and then combine it with recurrent consistency for training. However, due to the weak modeling capability of the model, their approach can only achieve redirection of simple motions and only supports the case of identical skeletal topologies. The graph convolution approach was proposed to achieve motion redirection between skeletons with different topologies. However, this approach has difficulty guaranteeing motion fidelity in any synthesized motions when faced with significant skeletal differences and complex movements [21]. Therefore, with the addition of deep learning, the technique is still worth exploring in the case of substantial skeletal differences and complex motions. For some sports with a short duration, such as golf swing, which lasts for about one second, the overall similarity between the user's action sequence and the standard action sequence can be directly compared to judge the completion quality of the action. However, for some sports items with longer duration, such as Taijiquan and radio gymnastics, which usually consist of multiple steps, if we only compare the overall similarity between the user's action sequence and the standard action sequence, the action evaluation is not accurate enough and cannot reflect the action details. For example, some activities are completed in a better manner for a set of Taijiquan movements completed by the user. Some are completed in a worse manner, and it is necessary to give a high score to make the activities conducted better.

3. Movement Evaluation Algorithm-Based Dancer Form Tracking Technology and Limb Optimal Control Model Construction

3.1. The Algorithmic Model for Movement Evaluation of Dancers Construction. Human movement evaluation is a technique to evaluate the completion quality of human movements. The system architecture includes four parts, namely, movement data acquisition, legal movement establishment, movement comparison and evaluation, and evaluation result output. Action data collection is mainly through the motion capture device Kinect to collect the coordinates of human skeletal nodes to obtain human action data [22]. The establishment of standard action is primarily through collecting professional coach's action data or through big data to get actionable data and, after the analysis of kinesiology experts, to establish the standard action template. Action comparison evaluation is done by calculating the error between user action sequences and common action sequences and evaluating the user's action concerning the error evaluation standard. The output of the evaluation result is to give the user a vivid image of the activity evaluation result by voice or image and to guide or score the user's action. The research of this paper focuses on the action comparison evaluation link. The architecture of the human action evaluation system is shown in Figure 1.

In the pre-evaluation section, the "action missing error" and "action sequence error" in the action sequence are judged to make a general evaluation of the action as a whole; in the action segmentation section, the action sequence is divided into subactions based on the action characteristics; in the detail evaluation section, the human action is evaluated according to the standardized one. In the action segment, the human activity is quantitatively analyzed according to the four evaluation indexes proposed in this study, namely, joint angle similarity, action center time similarity, action duration similarity, and combined average angular velocity similarity. The purpose of action pre-evaluation is to determine whether serious errors exist in the user's broadcast action sequence, i.e., "action missing errors" and "action sequence errors." The action pre-evaluation process consists of four main steps.

Let the user action sequence be $s = [P_1, P_2, \ldots, P_i, \ldots, P_n]$, and there are K standard poses, and let the kth classic pose be denoted as $B_k = (a_1, a_2, \ldots, a_8)$.

Calculate the Euclidean distance between the 1st standard pose B_1 and each element of the user action sequence $s = [P_1, P_2, \ldots, P_i, \ldots, P_n]$, where the Euclidean distance between the 1st classic pose B_1 and the 1st pose i $(i = 1, 2, \ldots, n)$ of the user action sequence p_i can be calculated by (1). Thus, the Euclidean distance vector between the 1st standard pose and the pose of the user action sequence B_1, n is obtained as $d = (d_1, d_2, \ldots, d_i, \ldots, d_n)$.

$$d_i = \sum \frac{b_1 - p_i}{\sqrt{b_1 + p_i}}. \tag{1}$$

In equation (1), B_1 is the joint angle vector of the 1st standard pose in the standard movement library; p_i is the joint angle vector of the user's i pose; D_i is the Euclidean distance between the user's ith pose and the 1st standard pose in the standard movement library.

(1) Search for the minimum value $d_{\min}$ in the elements of the Euclidean distance vector d; judge whether the minimum Euclidean distance $d_{\min}$ is less than the threshold O set in advance; if yes, proceed to the next step; if not, feedback to the user "action missing error" message.

(2) Repeat steps 1 and 2 for the other $k-1$ standard poses

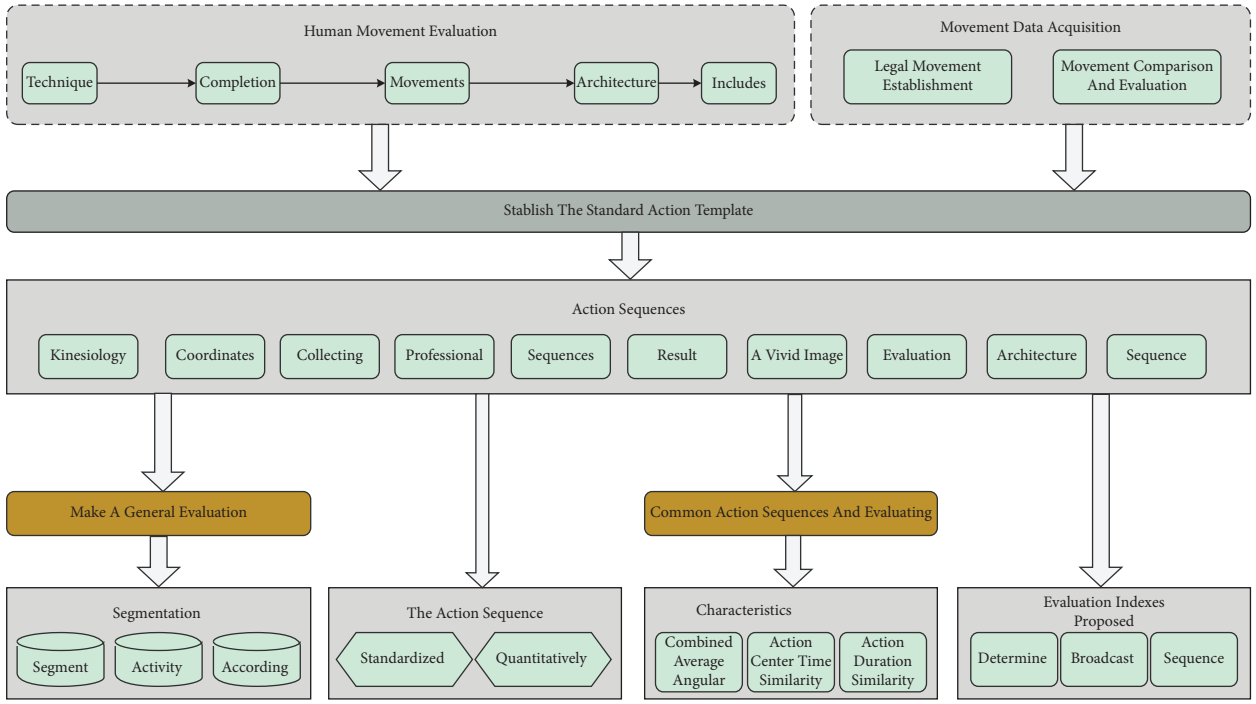

Figure 1: Architecture of the human movement evaluation system.

(3) Determine whether the order of K Euclidean distance minima in the user's action sequence is correct. If so, proceed to the action segmentation, and if not, give the user the "action order error" message.

Joint angle similarity includes static action joint angle similarity and dynamic action joint angle similarity, which are calculated by calculating the difference between each joint angle of the user and each joint angle of the standard action to measure the similarity between the user's movement and the legal action. Still, the static action joint angle and dynamic action joint angle similarity have different calculation methods. The static action joint angle similarity is calculated by calculating the average of all posture joint angles of the human household's static action and then comparing it with the average of all posture joint angles of the standard measure to obtain the similarity between them.

For the static action joint angle similarity d_s, let the action sequence of a static action of the user $C = (c_1, c_2, \ldots, c_n)$ and the standard pose vector corresponding to the static action be b. The static action joint angle similarity d_s is obtained by the following equation.

$$d_s = \sum \frac{\sqrt{c_1 + c_2 - c_n} - b}{n + b}. \tag{2}$$

In equation (2), c_i is the joint angle vector of i pose of the user's static action; b is the standard pose vector; n is the total number of frames of the user's static activity; d_s is the joint angle similarity/m of the user's static action.

For dynamic action joint angle similarity d_d, let the user's dynamic action sequence be $E = (e_1, e_2, \ldots, e_i, \ldots, e_n)$ and the standard action sequence corresponding to the dynamic action be $E = (e_1, e_2, \ldots, e_j, \ldots, e_m)$.

$$d = \int_{n=1} \frac{d_{11} - d_{12} + d_{1m}}{\sqrt{d_{n1} + d_{n2} - d_{nm}}} - d_{ij}. \tag{3}$$

The Euclidean distance between each poses composing the user dynamic action sequence and each pose containing the standard action sequence is calculated to obtain the distance matrix D, D_{ij} which denotes the Euclidean distance between the i posture of the user dynamic action and the j pose of the standard measure, which is obtained by the following equation.

$$d_{ij} = \sum \left(e_i + e_j\right)^2 + \left(e_i - e_j\right). \tag{4}$$

After obtaining the distance matrix D and the elements of the accumulation matrix g, the aspect g_{nm} in g can be calculated by the following recursive equation.

$$g = \sum_{n=1} \left(\frac{g_{11} - g_{12}}{g_{n1} + g_{n2} + g_{nm}} + g_{1m} \right), \tag{5}$$

$$g_{ij} = \sum \min \frac{g_{i-1} + g_{j-1}}{d_{ij}} - g_{i-1}. \tag{6}$$

The dynamic action joint angle similarity is calculated by equation (6) d_d. The g_{nm} size g_{nm} reflects the similarity between the user's dynamic action and its corresponding standard dynamic action. Still, to measure the completion quality between different activities in the user's long-time action behavior and form a unified evaluation index, this study gives the DTW distance g_{nm} divided by a factor, which is the sum of the total number of frames of the user's dynamic action n and its corresponding standard. This factor is the sum of the total number of frames m of the dynamic

action of the user and the total number of frames m of the corresponding standard action, which can reflect the DTW distance per unit frame of a dynamic movement of the user.

$$d_d = \sum g_{nm} \times \frac{\sqrt{n-m}}{\sqrt{n+m}}, \tag{7}$$

where n is the total number of frames of user's dynamic action; m is the total number of frames of standard dynamic action; g_{nm} is the shortest path from row 1 column 1 of the accumulation matrix G to row n column m of the accumulation matrix G, in degrees; d_d is the joint angle similarity of user's dynamic action.

To reflect the cumulative error in the time of user actions, the evaluation index used in this study is the action center time. It is known that f_{start} is the starting frame number of activities, f_{end} is the ending frame number of the action, and τ is the sampling frequency of the adopted motion capture device. The user's action center time t_c can be obtained by equation (8), and the similarity of action center time e_c can be obtained by equation (9), and t_c is the standard action center time.

$$t_c = \sum \frac{f_{\text{start}} + f_{\text{end}}}{t-2} + \frac{2}{t}, \tag{8}$$

$$e_c = \int (t_c - t) \times \left(t^2 - 1\right). \tag{9}$$

To reflect the error of individual joint angular velocity in dynamic action, the evaluation index used in this study is the average angular velocity similarity of dynamic action joints. The average angular velocity similarity of dynamic action joints is calculated as follows: (1) calculate the difference of joint angles between two adjacent frames; (2) calculate the average angular velocity of each joint; (3) calculate the Euclidean distance between the average angular velocity of user's dynamic action joints and the average angular velocity of standard dynamic action joints to obtain the average angular velocity similarity of dynamic action joints. Let the user's dynamic action sequence be $E = (e_1, e_2, \ldots, e_i, \ldots, e_n)$, and the average angular velocity of the standard joint of this dynamic action be w; the average angular velocity of the joint of this dynamic action can be calculated by the following equation.

$$e_w = \sum \frac{e_{i-1} + e_i}{n+1} + (t + w). \tag{10}$$

3.2. Dancers' Morphological Tracking Technology and Limb Optimal Control Model Design. The data acquisition module provides camera device connection, acquisition, and image data storage functions. After the hand dance process starts, the depth camera obtains data, aligns the acquired RGB map and depth map, and stores and passes the aligned image data to the hand action tracking and recognition module. In the hand action tracking and recognition module, hand segmentation is first performed after acquiring the image data to realize the separation of image foreground and background, and the hand in the image is extracted. Then, the segmented hand data is passed to the hand pose estimation algorithm to calculate the 3D coordinates of the joint hand points [23]. The 3D collective energy coordinate data are used for hand dance state data calculation in hand dance state data visualization on the one hand, and on the other hand, it will be used to calculate the features in the gesture recognition module to realize the recognition of the current gesture category. The gesture category recognition is used to determine whether the everyday gesture trained by the dancer is correct and to assist the dancer in practical hand dance training. In the hand dance state data visualization module, the hand rehabilitation state parameters are first calculated based on the hand 3D joint point coordinates. Then, the state data are visualized to the data changes with intuitive images. The flowchart of the data acquisition part is shown in Figure 2. After the camera is activated, it acquires the dancer's hand image data. After receiving the data, image alignment is performed for both types of images. The aligned images are stored on the one hand and transmitted to the hand movement tracking and recognition module for use on the other.

The hand motion tracking and the recognition module are divided into hand segmentation, hand pose estimation, and hand gesture recognition. After obtaining the RGB and depth maps in the data acquisition part, the hand segmentation algorithm is used to segment the hand from the image background. The implementation of hand segmentation can further improve the accuracy of hand gesture recognition and hand pose estimation. After hand segmentation, the hand image is passed into the hand gesture estimation method, and the corresponding algorithm is used to calculate the 3D joint point coordinates of the hand. After obtaining the hand collective point data, the hand common point data are transmitted to the hand gesture recognition module to judge whether the hand gesture is selected for training by the dancer and give corresponding hints in the image display interface. The hand joint point data are transmitted to the hand rehabilitation status data visualization module, which is used to calculate the hand dance status data.

The deep number of layers makes the feature map broader and more suitable for large datasets, and the network can solve 1000 classes of image classification and localization problems. The size of the convolutional kernel affects the number of parameters and the feeling field; the former relates to the difficulty of training and whether it is easy to deploy to mobile and so on. The latter relates to the update of parameters, the size of the feature map, whether the features are extracted enough, and the complexity of the model. The VGG convolutional neural network is the network structure of OpenPose to extract the information of human morphological features in images. The VGG convolutional neural network structure is divided into two branches and t stages [24]. One unit is used to detect the critical point heat map of the human body in the image, and the other branch is used to see the limb vector map of the human body in the picture. The output of each stage consists of a set S of key point heat maps and a human limb vector field L. The set S of key points contains J key point heat maps;

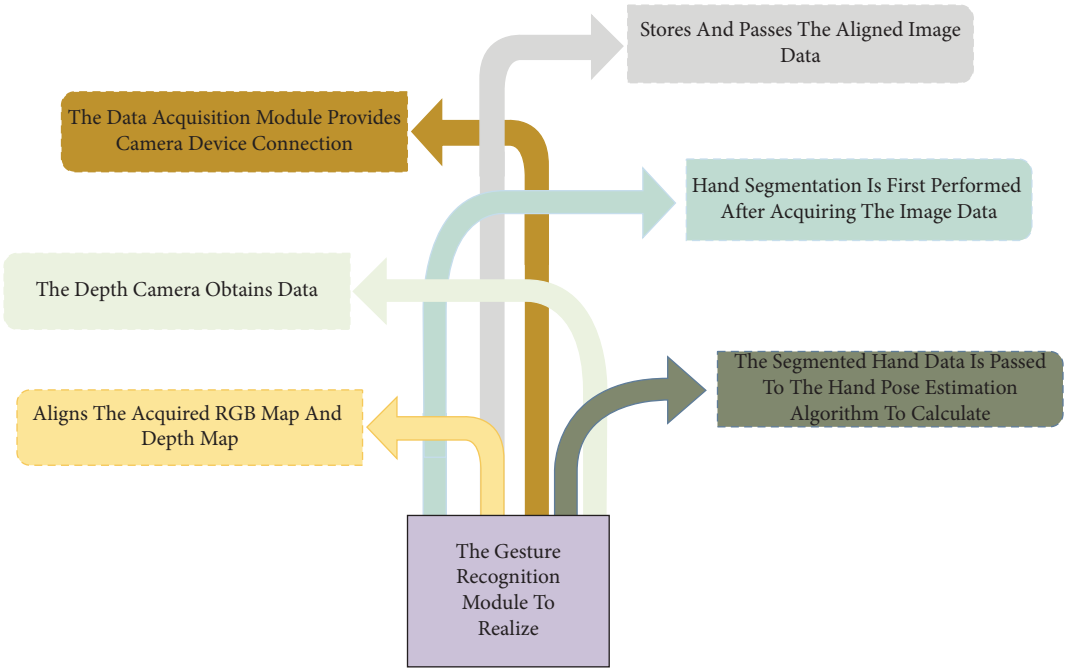

Figure 2: Flowchart of a data acquisition part.

each key point corresponds to one S_j; the human limb vector field L has C vector fields, and each limb corresponds to one L_c, the specific expressions are as follows:

$$S = \sum \frac{R^{w \times h} \times (J-1)}{s_1 - s_2 - s_J},$$
$$L = \frac{R^{w \times h \times 2} \times (c+L)^2}{L_1 - L_2 + L_c}. \tag{11}$$

For each of the t stages, the input is the crucial point heat map, the limb vector map of the previous step, and the original feature map. The output is the heat map of each key point of the human body and the vector map of the limb stem of each part of the human body. After iteration, the vital human points and human limb vectors are obtained, and the human morphology is constructed by evaluating the correlation between critical issues and connecting solid points. The specific evaluation method is that the dot product between the two key point connection vectors and the limb vectors of each pixel point on the two key point connection lines is calculated as the correlation between the two key points. Thus, the connection between the key points is performed to complete the human body morphology estimation.

Traditional OpenPose in dance movement morphology estimation is poor for movement morphology estimation. After observation and analysis, the motion amplitude and speed of upper and lower limbs are more significant, i.e., the interframe variation of key point positions is more significant, so the wrong estimation of key point positions and the wrong connection between critical points are easy to occur. Traditional OpenPose is only based on the images within a single frame in the video for human form estimation, and the interframe information of human form in the video is not used. Therefore, this study proposes an OpenPose method based on the morphological flow information between frames of human action in the video and an OpenPose method based on morphological flow optimization. By correcting the key point positions acquired by OpenPose through morphological flow, the interframe features of human morphology are used to improve the accuracy of human action morphology estimation. In this study, we offer the OpenPose method based on morphological flow optimization based on the traditional OpenPose utilizing the construction and solution methods of morphological flow [25]. The OpenPose method, based on morphological flow optimization, selects the intraframe critical point location information obtained by OpenPose to construct the interframe morphological flow information of human action and calculates the optimal solution of essential points of human morphology based on the morphological flow information to optimally correct the human morphology obtained based on OpenPose. The schematic diagram of the VGG convolutional neural network structure is shown in Figure 3.

This study estimates human morphology based on the OpenPose method of morphological flow optimization for Taijiquan movements. In the experimental results, the accuracy of the optimized OpenPose method is greatly improved. The OpenPose method based on morphological flow optimization is proposed. Based on the coordinate data of key points obtained by OpenPose, the human morphological flow information between frames is constructed. The optimal morphological solution is finally found, thus optimizing the human morphological estimation results obtained by OpenPose. Comparing the traditional OpenPose-based dance movement morphology estimation and the OpenPose-based morphology flow optimization, the proposed

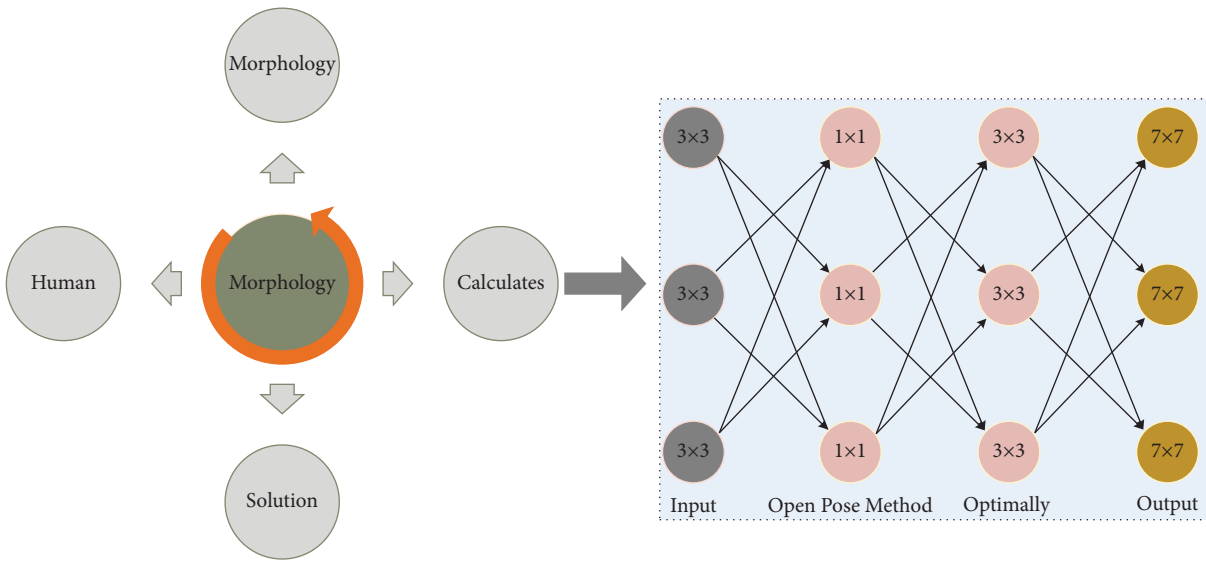

Figure 3: Schematic diagram of the convolutional neural network structure.

OpenPose-based morphology flow optimization method corrects and optimizes the positions of vital human points and effectively solves the problem of misconnection of human key points. In particular, the proposed OpenPose method significantly improves accuracy compared with the traditional OpenPose for estimating large amplitude and high-speed movements in Taijiquan, such as knee bending, golden chicken independence, and foot stomp.

4. Analysis of Results

4.1. Analysis of the Movement Evaluation Algorithm Model for Dancers. To verify the effectiveness of the proposed method for quantitative evaluation of complex human movements and the effectiveness of the improved DTW algorithm for improving the evaluation results, a movement evaluation comparison experiment based on the traditional DTW and the improved DTW algorithm was designed. In this study, considering the richness of the movements and the actual conditions, eight Taijis were selected as the evaluation objects. DTW is an algorithm used to compute the similarity of two time series. Unlike traditional methods that only calculate the Euclidean distance of elements at the same position, DTW allows one-to-many mapping. Each component of an input sequence Q can be mapped to the exact part of another sequence C and vice versa. However, the monotonicity of the series is guaranteed. DTW is generally implemented by dynamic programming. Eight experimenters were chosen to demonstrate dance movements in the establishment of the evaluation model; one was a high-level dancer. Four were used as the experimental objects for movement evaluation, and the rest were used as the data for establishing the score-distance mapping relationship. In this study, the dance moves of the player are the normal movement, corresponding to a score of 100 on a percentage scale, and the others, that is, the experimentalists have dance enthusiasts, novice dancers, and other players with different levels of dance movements. When each experimenter performs the dance movement demonstration, there is a professional score artificially, and the artificial score is used as the ideal algorithm scoring result.

In training the network, the number of layers and the number of neural units are adjusted to adjust the network structure and make the network converge and be in the best condition. The first half of the spatiotemporal cascade network is built based on Bi-LSTM and consists of 4 Bi-LSTM layers: 2 forward Bi-LSTM layers and 2 reverse Bi-LSTM layers. Each layer consists of 256 LSTM neurons, and the states of the neurons are randomly initialized. The connected ones are fully connected layers, the second half of the network is built based on CTC, and the network's output is the segmentation and recognition results. During the network training, the greedy search and beam search algorithms are compared to evaluate the change process of the loss function. The gradient descent is accelerated using the impulse optimization method Momentum Optimizer to optimize the convergence. The initial learning rate is adjusted to 0.01, the number of Bi-LSTM neurons, the batch size is 1, and the epoch is 150, so that the network converges to the minimum value and the network performance reaches the optimal state. The sample data score-distance fitting curve is shown in Figure 4.

The dataset used in this experiment is the dataset constructed in the motion evaluation algorithm. The dataset contains 6400 sequential motion capture data of different lengths for 16 lower limb movement categories. Each continuous movement data sample in the dataset includes 3–6 segments of a single type of movement; 16 movement types correspond to 16 Laban dance score symbols. All samples of this dataset were captured from different angles, thus ensuring sample diversity. Also, the experimental results will be more convincing and credible [26]. To reduce the dataset's complexity and the redundancy of the extracted spatial features, we remove some frames from each data sample in the dataset for subsequent processing. To ensure high data availability, the original boundaries of the data

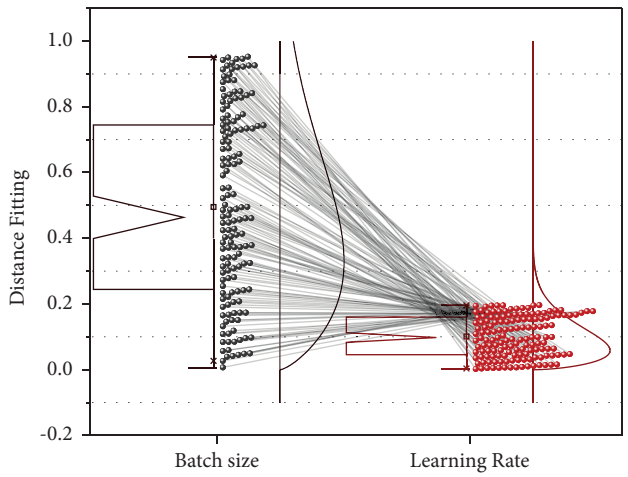

Figure 4: Sample data fraction-distance fitting curve.

samples are downsampled in this study, and the number of data frames is adjusted to 40 and 50 frames in two cases. It is found that the performance level can be maintained in the subsequent experiments when the number of data frames is 50 and 40. The performance remains the same when the number of data frames is 50 and 40. The network parameters are relatively less when the number of data frames is 30, which helps to reduce the network's training time. This process also saves the execution time of the continuous dance spectrum generation algorithm based on the spatio-temporal cascade network. The loss variation during network training is shown in Figure 5.

The attention mechanism-based segmentation method obtains the highest recognition accuracy on both datasets, while the unique segmentation-based approach has low accuracy. With the increase in the number of skeletal point delineation subsets, the recognition accuracy of both datasets improved significantly. Among them, the recognition accuracy of the spatial structure relationship delineation method is higher than that of the distance delineation method. It indicates that the number of skeletal delineation subsets affects the weighting of the weight function in the graph convolution. More skeletal delineation subset strategies benefit the graph convolution's hierarchical weighting of bony points. Therefore, the division strategy with more division subsets can obtain higher recognition accuracy. The attention mechanism division has the same skeletal subsets as the spatial structure relationship division. The attention mechanism division method incorporates the velocity characteristics of skeletons based on the spatial structure relationship of frames, i.e., the interframe displacement of the same skeletal points in the lean spatiotemporal map. Therefore, the attention mechanism-based segmentation method better simulates the strategy of human visual analysis, and the highest accuracy of human pose recognition is achieved in both datasets.

4.2. Movement Evaluation Algorithm-Based Dancer Form Tracking Technology and Limb Optimal Control Model

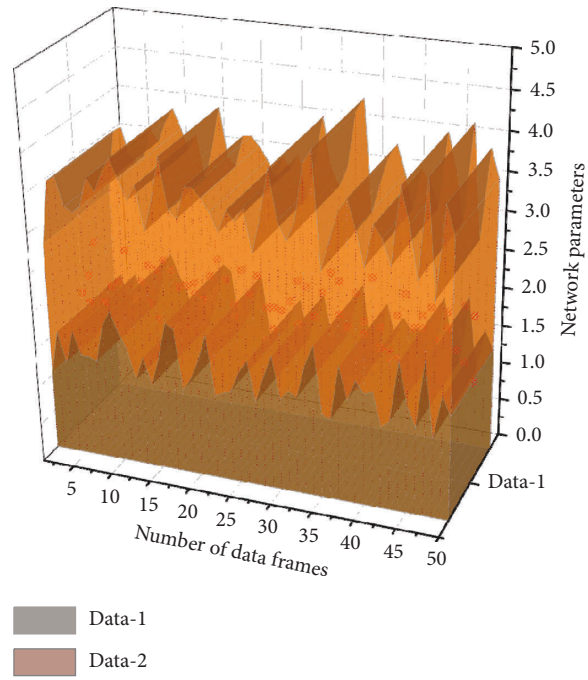

Figure 5: Change in loss during network training.

Implementation. After analysis, the present method achieves excellent performance in terms of accuracy and success rate. In addition, the current process has a significant speed advantage over the other two deep reinforcement learning-based methods, obtaining better results with an accuracy of 9.5% and 10.8% higher than the two candidate trackers, KCF, and STC, respectively, and it also brings an AUC value of 0.618. Since the intelligent body is only responsible for selecting the optimal candidate tracker in the proposed tracking framework, the target localization task is performed by the selected tracker. The experiments demonstrate the proposed approach's effectiveness, combining the advantages of different tracking algorithms to improve tracking accuracy. The evaluation results of the stochastic bits of intelligence cannot observe and analyze the tracking environment. This further indicates that the proposed method has learned the optimal strategy and has sufficiently mastered the advantages of the candidate trackers so that it can adaptively switch to the best tracker according to the current tracking environment. The frame-by-frame center localization errors in some test videos achieve long-term stable performance in target tracking by properly switching trackers. The above experimental results technique can improve the tracking performance by selecting the most appropriate tracker at different stages and successfully outputting more accurate positions as a priori knowledge for the next frame. The tracking results of the test set are compared, as shown in Figure 6.

In addition to tracking speed, tracking accuracy is another effort in monitoring research. To verify the impact of decision framework and specific decisions on tracking accuracy metrics, on the one hand, MACTFSS and MACTMTS are compared with Basic Tracker to independently verify the improvement of accuracy metrics by feature selection

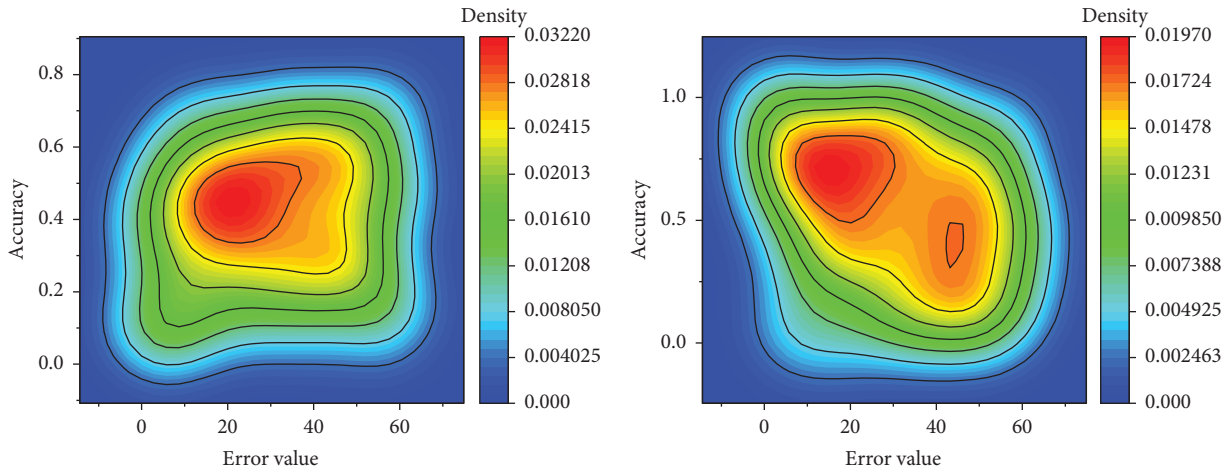

Figure 6: Comparison of trace results for the test set.

strategy FSS and motion trend strategy MTS, respectively; on the other hand, MACT is compared with MACTFSS and MACTMTS experiments to analyze whether the fusion of FSS and MTS has played an effect. The experiments demonstrate that the feature selection strategy FSS can improve the tracking speed; furthermore, it is hoped that this speed improvement should not reduce the accuracy requirement. The OPE success rate AUC score of MACTFSS is improved by 5.26%. The OPE accuracy AUC score is enhanced by 2.57% compared with the Basic Tracker. This suggests that the use of the feature selection strategy FSS not only did not reduce the tracking accuracy due to the use of fewer HOG features but also improved the tracking accuracy due to the flexible feature selection, which indirectly verified the correctness of the analysis of the observer attention change perspective. The test results of the optimal control model for morphological tracking limbs are shown in Figure 7.

The dance score generation algorithm based on spatial features takes full advantage of spatial feature fusion to improve the action recognition accuracy. The dance spectrum generation algorithm based on multitemporal modeling takes advantage of historical and future temporal information to enhance the impact on current prediction and optimize recognition accuracy. These two methods target single-movement recognition and cannot recognize continuous movements. The dance spectrum generation algorithm based on the spatiotemporal cascade network identifies ongoing actions, mainly to overcome the deficiency of needing to segment actions continuously; there is still room for improvement in performance optimization. On the one hand, because the predicted probability of the recognition network part takes a value less than 1, the total probability product is smaller than the expected probability of a single action.

On the other hand, there is an error in the category path merging when calculating the recognition accuracy of continuous actions resulting in the recognition accuracy of endless action categories is not as good as that of single steps. The joint network mechanism is currently the best regarding

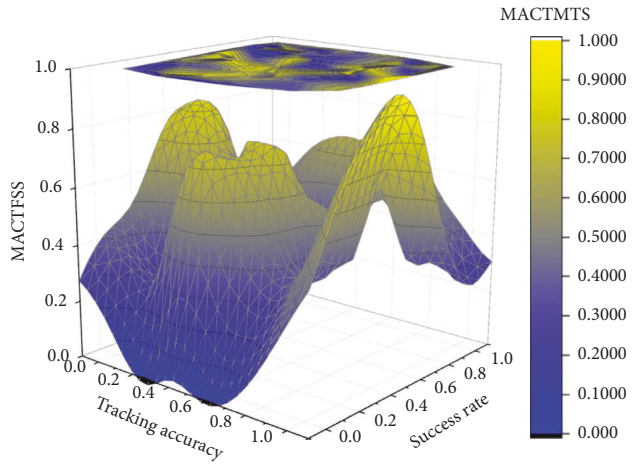

Figure 7: Test results of the optimal control model for morphological tracking limbs.

action recognition accuracy, but the network model is more complex, combining LSTM and CNN networks for model training. LSTM is memory, so the main feature of LSTM is that it has some memory capacity, so most of the time, it is used to deal with sequences, such as dealing with a sentence or a video. The CNN is mainly strong in dealing with a single picture, and the association between the front and back is not so strong in a sequence, but of course, the 3D CNN can also be used to deal with video some of the time. The method proposed in this study is relatively more straightforward and less computationally intensive than the joint network mechanism. It achieves long-time time series correlation modeling and spatiotemporal feature fusion, yielding good recognition accuracy. By comparing the experimental results, we can see that our approach has a performance improvement of about 2% in recognition accuracy compared to the joint network mechanism approach. When learners learn a movement, the captured action video sequence may be as many as a hundred frames; if they go frame by frame to compare the difference between their stance and the

standard stance, it takes a long time, learners may lose patience, and learning effect is not good, and second, there is no target to practice the stance one by one, which is less efficient; while if a swing action is decomposed, a few fundamental movements are selected and targeted to. If a swing is broken down and a few essential activities are selected for targeted practice, the stance at these basic movements will become closer and closer to the standard. The overall swing will become more and more standard, which is more efficient than correcting frame by frame.

5. Conclusion

In computer vision and image processing, the recognition of human behavior pose has become an important topic. It has been widely used in human-computer interaction, virtual reality, intelligent video surveillance, and so on. However, many problems still have not been well solved, which affect the computer's understanding and recognition accuracy of human behavior. This study proposes an OpenPose method based on pose flow optimization to address false detection of vital human points and misconnection between critical issues in traditional OpenPose-based human pose estimation. The optimal solution of the key points of the human pose is calculated based on the pose flow information between video frames to correct the human posture obtained by traditional OpenPose optimally. The OpenPose based on pose flow optimization compensates for the shortcoming of conventional OpenPose, which ignores the interframe information. The experimental results of the OpenPose method based on pose flow optimization improve the accuracy of human key point detection and effectively solve the problem of crucial point misconnection in the human pose. The mapping function is constructed by fusing the velocity characteristics and spatial structure relationship of the skeletal points in the neighborhood. The product of the velocity of bony points and the distance from the lean points to the body's center of gravity is used as the criterion to classify skeletal issues, i.e., the discriminant condition in the mapping function. The weight function in the graph convolution operation is improved based on the mapping function. Thus, the attention mechanism of human vision is simulated for human action analysis and recognition. The experimental results of the spatiotemporal map convolution network based on the attention mechanism can emphasize the action parts of human visual attention and improve the recognition accuracy compared with the traditional network model. In the action criteria evaluation, the linear regression method is used to model the extracted feature vectors. The DTW algorithm is used to match the curves of different lengths. The action evaluation experiment is designed, and a set of formulas are defined to evaluate the action based on the experimental data. The joint angle curve DTW difference is used as the experimental parameter, and the rationality of the activity evaluation method is demonstrated through the action evaluation experiment. In the study of keyframe acquisition based on human pose estimation and clustering, it was found that the use of the fixed clustering center approach to obtain keyframes leads to the fact that even if there are repeated actions in the video, only one of the closest results to the standard frame but not to multiple keyframes, and the single development does not. Therefore, in the future research, we will consider segmenting the video according to specific information to determine the position of the repetitive frames. Therefore, in future research, we will consider segmenting the footage based on detailed information to determine that there are no repetitive actions in each segment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the College of Music and Dance, Yulin Normal University.

References

[1] L. Xin, "Evaluation of factors affecting dance training effects based on reinforcement learning," *Neural Computing & Applications*, vol. 34, no. 9, pp. 6773–6785, 2022.

[2] W. Xu, A. Chatterjee, M. Zollhofer et al., "Mo^{2}Cap^2: real-time mobile 3D motion capture with a cap-mounted fisheye camera," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 2093–2101, 2019.

[3] Z. Yu, X. Shi, J. Zhou et al., "Feasibility of the indirect determination of blast-induced rock movement based on three new hybrid intelligent models," *Engineering with Computers*, vol. 37, no. 2, pp. 991–1006, 2021.

[4] C. Yuanyuan, W. Rui, Z. Bin, and W. S. Griffith, "Temporal association rules discovery algorithm based on improved index tree," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 115–128, 2021.

[5] Y. Wei and J. Zhao, "Designing human-like behaviors for anthropomorphic arm in humanoid robot NAO," *Robotica*, vol. 38, no. 7, pp. 1205–1226, 2020.

[6] S. Yanfei, Z. Yahong, W. Tianxiong, and Z. Lin, "An algorithm of moving pieces to become black alternation with white based on dimension reduction," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 163–170, 2021.

[7] M. Zhou, H. Dong, P. A. Ioannou, Y. Zhao, and F. Y. Wang, "Guided crowd evacuation: approaches and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 5, pp. 1081–1094, 2019.

[8] B. Wang, "Attribute reduction method based on sample extraction and priority," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 219–226, 2021.

[9] T. A. Sulaiman, H. Bulut, and S. S. Atas, "Optical solitons to the fractional Schrödinger-Hirota equation," *Applied Mathematics and Nonlinear Sciences*, vol. 4, no. 2, pp. 535–542, 2019.

[10] H. Yokota, M. Naito, N. Mizuno, and S. Ohshima, "Framework for visual-feedback training based on a modified self-organizing map to imitate complex motion," *Proceedings of the Institution of Mechanical Engineers - Part P: Journal of*

Sports Engineering and Technology, vol. 234, no. 1, pp. 49–58, 2020.

[11] M. R. Keyvanpour, S. Vahidian, and M. Ramezani, "HMR-vid: a comparative analytical survey on human motion recognition in video data," *Multimedia Tools and Applications*, vol. 79, no. 43-44, Article ID 31863, 2020.

[12] N. Ding and H. Guo, "Energy-saving design of office buildings considering light environment and thermal environment," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 269–282, 2021.

[13] W. Zhao, T. Shi, and L. Wang, "Fault diagnosis and prognosis of bearing based on hidden markov model with multi-features," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 1, pp. 71–84, 2020.

[14] G. Jianbang and S. Changxin, "Real-time monitoring of physical education classroom in colleges and universities based on open IoT and cloud computing," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 4, pp. 7397–7409, 2021.

[15] W. Hao, D. Rui, L. Song, Y. Ruixiang, Z. Jinhai, and C. Juan, "Data processing method of noise logging based on cubic spline interpolation," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 93–102, 2021.

[16] H. Luo, X. Hu, Y. Zou, X. Jing, C. Song, and Q. Ni, "Research on a reference signal optimisation algorithm for indoor Bluetooth positioning," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 2, pp. 525–534, 2021.

[17] X. Qi, H. Li, B. Chen, and G. Altenbek, "A prediction model of urban counterterrorism based on stochastic strategy," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 263–268, 2021.

[18] Y. Wu, Z. Huo, W. Xing, Z. Ma, and H. M. A. Ahmed, "Application of experience economy and recommendation algorithm in tourism reuse of industrial wasteland," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 2, pp. 227–238, 2021.

[19] M. Wu, A. Payshanbiev, Q. Zhao, and W. Yang, "Nonlinear optimization generating the tomb mural blocks by GANS," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 43–56, 2021.

[20] K. Yan, S. Jinling, B. Mingming, F. Haipeng, and M. Salama, "Red tide monitoring method in coastal waters of Hebei Province based on decision tree classification," *Applied Mathematics and Nonlinear Sciences*, vol. 7, no. 1, pp. 43–60, 2022.

[21] S. Debbarma and S. Bhadra, "A lightweight flexible wireless electrooculogram monitoring system with printed gold electrodes," *IEEE Sensors Journal*, vol. 21, no. 18, Article ID 20942, 2021.

[22] Y. Lin, S. Li, K. Jia, and K. L. Kingsley, "The research of power allocation algorithm with lower computational complexity for non-orthogonal multiple access," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 2, pp. 79–88, 2021.

[23] Q. Zhang, "Fully discrete convergence analysis of non-linear hyperbolic equations based on finite element analysis," *Applied Mathematics and Nonlinear Sciences*, vol. 4, no. 2, pp. 433–444, 2019.

[24] M. Zhuang, H. Li, and Y. Lin, "A novel joint transmitting and receiving antenna selection for spatial multiplexing systems," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 2, pp. 565–580, 2020.

[25] H. Jingchao and H. Zhang, "Recognition of classroom student state features based on deep learning algorithms and machine learning," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2361–2372, 2021.

[26] L. Liu, M. Niu, D. Zhang, L. Liu, and D. Frank, "Optimal allocation of microgrid using a differential multi-agent multi-objective evolution algorithm," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 2, pp. 111–124, 2021.